arXiv:1304.7700v2 [physics.comp-ph] 30 Apr 2013

INFORMATION-THEORETIC TOOLS FOR PARAMETRIZED COARSE-GRAINING OF NON-EQUILIBRIUM EXTENDED SYSTEMS*

MARKOS A. KATSOULAKIS† AND PETR PLECHÁČ‡

Abstract. In this paper we focus on the development of new methods suitable for efficient and reliable coarse-graining of *non-equilibrium* molecular systems. In this context, we propose error estimation and controlled-fidelity model reduction methods based on Path-Space Information Theory, and combine it with statistical parametric estimation of rates for non-equilibrium stationary processes. The approach we propose extends the applicability of existing information-based methods for deriving parametrized coarse-grained models to Non-Equilibrium systems with Stationary States (NESS). In the context of coarse-graining it allows for constructing optimal parametrized Markovian coarse-grained dynamics, by minimizing information loss (due to coarse-graining) on the path space. Furthermore, the associated path-space Fisher Information Matrix can provide confidence intervals for the corresponding parameter estimators. We demonstrate the proposed coarse-graining method in a non-equilibrium system with diffusing interacting particles, driven by out-of-equilibrium boundary conditions.

Key words. coarse-grained dynamics, non-equlibrium stationary states, driven difussion, relative entropy rate, Fisher information matrix, parametrization, kinetic Monte Carlo, Markov processes

1. Introduction. Non-equilibrium systems at transient or steady state regimes are typical in applied science and engineering, and are the result of coupling between different physicochemical mechanisms, driven by external couplings or boundary conditions. Typical examples include reaction-diffusion systems in heteroepitaxial catalytic materials, polymeric flows and separation processes in microporous materials, [29, 25, 28]. In this paper we develop reliable model-reduction methods capable to handle extended, *non-equilibrium* statistical mechanics models and related multi-physics systems. Model-reduction (or coarse-graining) approaches can be often described in the context of parameter estimation of parametrized statistical models. However, atomistic models of materials lead to high-dimensional probability distributions and/or stochastic processes to which the standard methods of statistical inference and model discrimination are not directly applicable. The emphasis on information theory tools is also partly justified since often we are interested in probability density functions (PDF), typically non-Gaussian, due to the significance of tail events in complex systems. A primary focus of this paper is on systems with Non-Equilibrium Steady States (NESS), i.e., systems in which a steady state is reached but the detailed balance condition is violated and explicit formulas for the stationary distribution, e.g., in the form of a Boltzmann distribution, are not available.

Application of information-theoretic methods to analysis of stochastic models uses entropy-based techniques for analyzing and estimating a distance between (probability) measures. The relative entropy (Kullback -Leibler divergence) of two probability measures $\mu(dx) = \mu(x)\,dx$ and $\nu(dx) = \nu(x)\,dx$

$$\mathcal{R}\left(\mu \,|\, \nu\right) = \int \mu(x) \log \frac{\mu(x)}{\nu(x)}\, dx$$

*Submitted to the Journal of Chemical Physics.

†Department of Mathematics and Statistics, University of Massachusetts, Amherst, MA 01003,markos@math.umass.edu

‡Department of Mathematical Sciences, University of Delaware, Newark, DE 19716,plechac@math.udel.edu

allows us to define a pseudo-distance between two measures. A key property of the relative entropy $\mathcal{R}(P\,|\,Q)$ is that $\mathcal{R}(P\,|\,Q) \geq 0$ with equality if and only if $P = Q$, which allows us to view relative entropy as a "distance" (more precisely a semi-metric) between two probability measures P and Q. Moreover, from an information theory perspective [7], the relative entropy measures *loss/change of information*. Relative entropy for high-dimensional systems was used as measure of loss of information in coarse-graining [18, 14, 2], and sensitivity analysis for climate modeling problems [22].

Using entropy-based analytical tools has proved essential for deriving rigorous results for passage from interacting particle models to mean-field description, [20]. The application of relative entropy methods to the error analysis of coarse-graining of stochastic particle systems have been introduced and studied in [14, 16, 17, 15, 19]. Aside of this rigorous numerical analysis direction, entropy-based computational techniques were also developed which are used for constructing approximations of coarse-grained (effective) potentials for models of large biomolecules and polymeric systems (fluids, melts). Optimal parametrization of effective potentials based on minimizing the relative entropy between *equilibrium* Gibbs states, e.g.,[6, 5, 3], extended previously developed inverse Monte Carlo methods, primarily based on force matching approaches, used in coarse-graining of macromolecules (see, e.g., [31, 24]). In [10] an extension to dynamics is proposed in the context of Fokker-Planck equations, by considering the corresponding relative entropy for discrete-time approximations of the transition probabilities. Furthermore, relative entropy was used as means to improve model fidelity in a parametric, multi-model approximation framework of complex dynamical systems, at least when the model's steady-state distributions are explicitly known, e.g. are Gaussian, [23].

Overall, such parametrization techniques are focusing on systems with a known steady state, such as a Gibbs equilibrium distribution. More specifically, computational implementations of optimal parametrization in the inverse Monte Carlo methods is relatively straightforward for equilibrium systems in which the best-fit procedure is applied to an *explicitly known* equilibrium distribution and where relative entropy is explicitly computable. On the other hand, this is not the case in non-equilibrium systems, even at a steady state where typically we do not have a Gibbs structure and the steady-state distribution is unknown altogether, setting up one of the primary challenges for this paper. Indeed, here we show that, in non-equilibrium systems the general *information theory* ideas based on Kullback-Leibler divergence are still applicable but they have to be properly formulated in the context of Non-equilibrium Statistical Mechanics by focusing on the probability distribution of the entire time series, i.e. on the *path space* of the underlying stochastic processes. We show that, surprisingly, such a path-space relative entropy formulation is: (a) general in the sense that it applies to any Markovian models (e.g. Langevin dynamics, Kinetic Monte Carlo, etc), and (b) is easily computable as an ergodic average in terms of the *Relative Entropy Rate*, therefore allowing us to construct optimal parametrized Markovian coarse-grained dynamics for large classes of models. This procedure involves the minimization of information loss in path space, where information is inadvertently lost due to coarse-graining procedure. In fact, the proposed parametrization scheme in [10] is mathematically justified by reformulating it on the path space using the Relative Entropy Rate and it is a specific, but reversible (i.e. it has a Gibbs steady state) example of our methodology. Finally, the path-space Fisher Information Matrix (FIM) derived from Relative Entropy Rate (RER) can provide confidence intervals for the corresponding statistical estimators of the optimal parameter obtained through the

minimization problem.

The paper is structured as follows. In Section 2 we formulate the path-space information theory tools used in the paper, including the key concept of Relative Entropy Rate. In Section 3 we present the parameterization method of coarse-grained dynamics and the connections with maximum likelihood estimators and the Fisher Information Matrix. In Section 4 we briefly discuss statistical estimators for RER and FIM. Finally, in Section 5 we demonstrate the proposed coarse-graining method in a non-equilibrium system with diffusing interacting particles, driven by out-of-equilibrium boundary conditions.

2. Relative Entropy Rate, path space information theory and error quantification for non-equilibrium systems. First, we formulate a general entropy -based error analysis for coarse-graining, dimensional reduction and parametrization of high-dimensional Markov processes, simulated by Kinetic Monte Carlo (KMC) and Langevin Dynamics. Typically such systems have Non-Equilibrium Steady States (NESS) for which *detailed balance fails* as they are *irreversible*. The stationary distributions are not known explicitly and has to be studied computationally. Quantifying and controlling the coarse-graining error in such systems thus requires to develop *computable and efficient* methods for estimating distances of probability measures on the path space. The relative entropy between two path measures $P_{[0,T]}$ and $Q_{[0,T]}$ (see (2.3) for a specific example) for the processes on the interval $[0,T]$ is

$$\mathcal{R}\left(P_{[0,T]} \,|\, Q_{[0,T]}\right) = \mathbb{E}_{P_{[0,T]}}\left[\log \frac{dP_{[0,T]}}{dQ_{[0,T]}}\right], \tag{2.1}$$

where $\frac{dP_{[0,T]}}{dQ_{[0,T]}}$ is the Radon-Nikodym derivative of $P_{[0,T]}$ with respect to $Q_{[0,T]}$. If these probability measures have probability densities p, q respectively, (2.1) becomes $\mathcal{R}\left(P_{[0,T]} \,|\, Q_{[0,T]}\right) = \int p \log\left(\frac{p}{q}\right)$. In the setting of coarse-graining or model-reduction the measure $P_{[0,T]}$ is associated with the exact process and $Q_{[0,T]}$ with the approximating (coarse-grained) process.

From an information theory perspective, the relative entropy measures the loss of information as we approximate the exact stochastic process $P_{[0,T]}$ with the coarse-grained one $Q_{[0,T]}$. In general the relative entropy (2.1) in this dynamic setting is not a computable object; we refer for instance to related formulas in the Shannon-MacMillan-Breiman Theorem, [7]. However, as we show next, in practically relevant cases of *stationary* Markov processes we can work with the *relative entropy rate*

$$\mathcal{H}(P \,|\, Q) = \lim_{T\to\infty} \frac{1}{T}\mathcal{R}\left(P_{[0,T]} \,|\, Q_{[0,T]}\right), \tag{2.2}$$

where P and Q denote the distributions of the corresponding stationary processes.

Relative Entropy Rate for Markov Chains. In order to explain the basic concept we restrict to the case of two Markov chains, $\{X_n\}_{n\geq 0}$, $\{\tilde{X}_n\}_{n\geq 0}$ on the countable state space Σ, defined by the transition probability kernels $p(x,x')$ and $q(x,x')$. A typical example would be the embedded Markov chain used for KMC simulations of a continuous time Markov chain. In the case of a continuous state space a temporal discretization of a Langevin process, leads to a Markov process with the transition kernel $p(x,dx') = p_{\Delta t}(x,x')\,dx'$ defined by the time-discretization scheme of the underlying stochastic dynamics. We assume that the initial states are from the invariant distributions $\mu(x)$ and $\nu(x)$. The path measure defining the probability of a path

$(x_0, x_1, \ldots, x_T)$ is then

$$P(x_0, \ldots, x_T) = \mu(x_0)p(x_0, x_1)\ldots p(x_{T-1}, x_T)\,, \tag{2.3}$$

and similarly for the measure $Q(x_0, \ldots, x_T)$. The Radon-Nikodym derivative is easily computed

$$\frac{dP}{dQ} = \frac{\mu(x_0)\prod_{i=0}^{T-1} p(x_i, x_{i+1})}{\nu(x_0)\prod_{i=0}^{T-1} q(x_i, x_{i+1})}\,.$$

Using the fact that the processes are stationary with invariant measures μ and ν, we obtain an expression for the relative entropy

$$\mathcal{R}\left(P\,|\,Q\right) = T\,\mathbb{E}_\mu\left[\sum_{x'\in\Sigma} p(x,x')\log\frac{p(x,x')}{q(x,x')}\right] + \mathcal{R}\left(\mu\,|\,\nu\right)\,, \tag{2.4}$$

and thus the relative entropy rate is given explicitly as

$$\mathcal{H}(P\,|\,Q) = \sum_{x\in\Sigma}\mu(x)\sum_{x'\in\Sigma} p(x,x')\log\frac{p(x,x')}{q(x,x')}]\,. \tag{2.5}$$

We will refer from now on to the quantity (2.5) as the *Relative Entropy Rate* (RER), which can be thought as the change in information per unit time. Notice that RER has the correct time scaling since it is actually independent of the interval $[0, T]$. Furthermore, it has the following key features that make it a crucial observable for simulating and coarse-graining complex dynamics:

(i) The RER formula (2.5) provides a computable observable that can be sampled from the steady state μ in terms of conventional Kinetic Monte Carlo (KMC), bypassing the need for a histogram or an explicit formula for the high-dimensional probabilities involved in (2.1).

(ii) In stationary regimes, when $T \gg 1$ in (2.4), the term $\mathcal{R}\left(\mu\,|\,\nu\right)$ becomes unimportant. This is especially convenient since μ and ν are typically not known explicitly in non-reversible systems, for instance in reaction-diffusion or driven-diffusion KMC or non-reversible Langevin dynamics.

In view of these features, we readily see that if we consider a Markov chain $\{\tilde{X}_n\}_{n\geq 0}$ as an approximation, e.g. a coarse-graining, of the chain $\{X_n\}_{n\geq 0}$, we can estimate the loss of information at long times by computing $\mathcal{H}(P\,|\,\tilde{P})$ as an ergodic average. This observation is the starting point of the proposed methodology and relies on the fact that the observable $\mathcal{H}(P\,|\,Q)$ is *computable*; efficient statistical estimators for (2.5) are discussed in Section 4. A similar calculation can be carried out for continuous time Markov Chains, as we see next.

Continuous Time Markov chains and Kinetic Monte Carlo. In models of catalytic reactions the systems are often described by continuous time Markov chains (CTMC) that are simulated by KMC algorithms. For example, the microscopic Markov process $\{\sigma_t\}_{t\geq 0}$ describes the evolution of molecules on a substrate lattice. Mathematically the continuous time Markov chain is defined completely by specifying the local transition rates $c^\theta(\sigma, \sigma')$ where $\theta \in \mathbb{R}^k$ is a vector of the model parameters. The transition rates determine the updates from any current state (configuration) $\sigma_t = \sigma$ to a (random) new state σ'. In the context of the spatial models considered here, the transition rates take the form $c^\theta(\sigma, \sigma') = c^\theta(x, \omega, \sigma)$, denoting by $x \in \Lambda_N$ a lattice site on a d-dimensional lattice Λ_N and $\omega \in \mathcal{S}_{Nx}$, where $\mathcal{S}_{Nx}$ is the set of all possible configurations

that correspond to an update in a neighborhood of the site x. From local transition rates one defines the total rate $\lambda^\theta(\sigma) = \sum_{x\in\Lambda_N}\sum_{\omega\in\mathcal{S}_{N_x}} c^\theta(x,\omega,\sigma)$, which is the intensity of the exponential waiting time for a jump from the state σ. The transition probabilities for the embedded Markov chain $\{S_n\}_{n\geq 0}$ are $p(\sigma,\sigma';\theta) = \frac{c(x,\omega,\sigma;\theta)}{\lambda(\sigma;\theta)}$. In other words once the exponential "clock" signals a jump, the system transitions from the state σ to a new configuration σ' with the probability $p(\sigma,\sigma')$. In the context of coarse-graining or hybrid systems we are led to finding an optimal parametrization for the rates $\tilde{c}(\sigma,\sigma';\theta)$ of a processes that approximates the dynamics given by the microscopic process $c(\sigma,\sigma')$. A similar calculation as in the case of Markov chains gives the analogue of the formula (2.5)

$$\mathcal{H}(P\,|\,Q) = \mathbb{E}_\mu\left[\lambda(\sigma) - \tilde{\lambda}(\sigma;\theta) - \sum_{\sigma'} c(\sigma,\sigma')\log\frac{c(\sigma,\sigma')}{\tilde{c}(\sigma,\sigma';\theta)}\right], \tag{2.6}$$

where μ is the stationary distribution of the microscopic process and λ denotes total transition rates. In [13] we used this quantity in order to quantify error in a two-level coarse-grained kinetic Monte Carlo method. Based on these considerations, we show in Section 3 that minimizing the error measured by (2.6) leads to a Markovian coarse-grained dynamics that best approximates long-time behavior of the microscopic process projected to the coarse degrees of freedom.

Remark 2.1. We consider the special case where the transition probability function of the Markov chain is sampled directly from the invariant measure, i.e.

$$p(\sigma,\sigma') = \mu(\sigma'), \text{ and } q(\sigma,\sigma') = \nu(\sigma'), \text{ for all } \sigma,\sigma'\in\Sigma.$$

This sampling is equivalent to the fact that the path space samples in (2.4) are independent and identically distributed from the stationary probability distributions. Then the RER between the path probabilities becomes the usual relative entropy between the stationary distributions:

$$\mathcal{H}(P\,|\,Q) = \mathcal{R}\left(\mu\,|\,\nu\right). \tag{2.7}$$

Estimating RER using (2.5) is far simpler than directly estimating the relative entropy $\mathcal{R}\left(\mu\,|\,\nu\right)$, since (2.5) only involves local dynamics rather than the full steady state measure, which typically may not be available. Furthermore even when it is available in the form of a Gibbs state it will require computations that will typically involve a full Hamiltonian, [6].

Estimation and error of observables. The estimates on relative entropy and RER can provide an upper bound for a large family of observable functions through the Pinsker's (or Csiszar-Kullback-Pinsker) inequality. The Pinsker inequality states that the total variation norm between $P_{[0,T]}$ and $Q_{[0,T]}$ is bounded in terms of the relative entropy, [7]. The Pinsker inequality gives an estimate for a difference of the mean computed with respect to the distribution P and Q

$$|\mathbb{E}_{P_{[0,T]}}[f] - \mathbb{E}_{Q_{[0,T]}}[f]| \leq ||f||_\infty\sqrt{2\mathcal{R}\left(P_{[0,T]}\,|\,Q_{[0,T]}\right)}. \tag{2.8}$$

An important conclusion that is immediately drawn from the above inequality is that if the relative entropy of a distribution with respect to another distribution is small then the error between any bounded observables function is also accordingly small. Using (2.4) we readily obtain the estimate

$$|\mathbb{E}_{P_{[0,T]}}[f] - \mathbb{E}_{Q_{[0,T]}}[f]| \leq ||f||_\infty\sqrt{2T}\sqrt{\mathcal{H}(P\,|\,Q) + \frac{1}{T}\mathcal{R}\left(\mu\,|\,\nu\right)}, \tag{2.9}$$

involving the relative entropy rate (2.5) or (2.6). As in virtually all numerical analysis estimates for stochastic dynamical systems, the bound (2.9) may not be sharp, but it is indicative of the error in the observables when the distribution Q approximates P.

3. Parametrization of coarse-grained dynamics and Inverse Dynamic Monte Carlo. We consider a parameterized class of coarse-grained Markov processes $\{\eta_t\}_{t\geq 0}$, associated with the fine scale stochastic process $\{\sigma_t\}_{t\geq 0}$. The coarse-graining procedure is based on projecting the microscopic space Σ into a coarse space $\bar{\Sigma}$ with less degrees of freedom. We denote the coarse space variables

$$\eta = \mathbf{T}\sigma\,, \quad \text{where} \quad \mathbf{T} : \Sigma \to \bar{\Sigma} \tag{3.1}$$

is a coarse-graining (projection) operator, see also (5.2) for a specific example. In the case of continuous time processes such as Kinetic Monte Carlo, the coarse-grained stochastic process is defined in terms of coarse transition rates $\bar{c}(\eta, \eta')$ which captures macroscopic information from the fine scale rates $c(\sigma, \sigma')$. For example, for stochastic lattice systems, approximate coarse rate functions are explicitly known from coarse graining (CG) techniques of [15, 19], see (5.1). Similarly, when we consider temporally discretized stochastic processes such as Langevin Dynamics, the coarse-grained process is given in terms of transition probabilities $\bar{p}(\eta, \eta')$ which capture macroscopic information from the fine scale transition probabilities $p(\sigma, \sigma')$.

Interpolated dynamics. Given coarse-grained dynamics we can always construct corresponding microscopic dynamics. For example, given coarse-grained transition probabilities $\bar{p}(\eta, \eta')$ with corresponding stationary distribution $\bar{\mu}$, where the latter is typically unknown in non-equilibrium systems, we define the corresponding fine-scale rates

$$q(\sigma, \sigma') := \bar{p}(\mathbf{T}\sigma, \mathbf{T}\sigma')\,. \tag{3.2}$$

Here we apply piece-wise constant interpolation for all microscopic states σ (reps. σ') corresponding to the same coarse state η (reps. η') and thus transitions to these states occur with the same probability rates. The reconstruction step (3.2) is necessary when we want to compare fine and coarse processes on the path space in terms of the relative entropy rate, since both processes need to be defined on the same probability space, see for example (3.5) below. In this paper, for the sake of simplicity, we assume that all reconstructions are based on (3.2). The reconstruction is obviously not unique and a complete discussion and related strategies are given in [30].

3.1. Inverse Dynamic Monte Carlo methods.. In many applications the coarse-grained models are defined by effective potentials or effective rates which are sought in a family of parameter-dependent functions, [31, 24, 21]. The parameters are then fitted by minimizing certain functionals that attempt to capture different aspects of modeling errors, e.g., radial distribution functions in [24]. Compared to such Inverse Monte Carlo methods applied to equilibrium systems we cannot work with equilibrium distributions since the primary information about the non-equilibrium system is represented by the transition rates and the NESS is not known. Thus we apply the information-theoretic framework on the path space, i.e., the approximating measure $Q_{[0,T]} \equiv Q^{\theta}_{[0,T]}$ depends on the parameters $\theta \in \mathbb{R}^k$ that are fitted using entropy based criteria for the best approximation.

The optimal parametrized coarse-grained transition probabilities $q^{\theta^*}(\sigma, \sigma')$ are constructed as follows. First, given the parametrized coarse-grained transition probabilities $\bar{p}^{\theta}(\eta, \eta')$ we define the fine-scale projected rates $q^{\theta}(\sigma, \sigma')$, which can be defined,

for instance, by (3.2) as

$$q^\theta(\sigma,\sigma') = \bar{p}^\theta(\mathbf{T}\sigma,\mathbf{T}\sigma')\,,$$

and the corresponding coarse-grained path-distribution is

$$Q^\theta(\sigma_0,\dots,\sigma_T) = \bar{\mu}(\mathbf{T}\sigma_0)q^\theta(\sigma_0,\sigma_1)\dots q^\theta(\sigma_{T-1},\sigma_T)\,. \tag{3.3}$$

Subsequently the best-fit can be obtained by minimizing the relative entropy rate, i.e., finding solution

$$\theta^* = \arg\min_\theta \mathcal{H}(P\,|\,Q^\theta)\,, \tag{3.4}$$

where now we have that the RER is

$$\mathcal{H}(P\,|\,Q^\theta) = \sum_{\sigma\in\Sigma}\mu(\sigma)\sum_{\sigma'\in\Sigma}p(\sigma,\sigma')\log\frac{p(\sigma,\sigma')}{q^\theta(\sigma,\sigma')}]\,. \tag{3.5}$$

This optimization problem on one hand is similar to more common parametric inference in which the log-likelihood function is maximized, and this perspective will be further clarified in Section 3.2. Furthermore, due to the parametric identification of the coarse-grained dynamics, i.e., transition probabilities, or rates in the case of (2.6), we refer to the proposed methodology as an *Inverse Dynamic Monte Carlo* method in analogy to the Inverse Monte Carlo methods for equilibrium systems, [31, 24, 21].

The optimization algorithm for (3.4) is based on iterative procedures that locate a solution θ^* of the optimality condition $\nabla_\theta\mathcal{H}(P\,|\,Q^\theta)=0$ of the following type

$$\theta^{(n+1)} = \theta^{(n)} - \frac{\alpha}{n}G^{(n+1)}\,, \tag{3.6}$$

for some $\alpha>0$ and $G^{(n+1)}$ being a suitable approximation of the gradient $\nabla_\theta\mathcal{H}(P\,|\,Q^\theta)$, more precisely $\mathbb{E}[G^{(n+1)}\,|\,G^{(0)},\theta^{(0)},\dots,G^{(n)},\theta^{(n)}] = \nabla_\theta\mathcal{H}(P\,|\,Q^\theta)$. The crucial ingredient of this algorithm is an efficient and reliable estimator for the sequence $G^{(n)}$ of the gradient estimates. Similar to the deterministic case the minimization can be accelerated by combining this step with the Newton-Raphson method and choosing the vector G as

$$G^n = \mathrm{Hess}(\mathcal{H}(P\,|\,Q^{\theta^n}))^{-1}\nabla_\theta\mathcal{H}(P\,|\,Q^{\theta^n})\,. \tag{3.7}$$

While the evaluation of the Hessian $\mathrm{Hess}(\mathcal{H}(P\,|\,Q^{\theta^n}))$ presents an additional computational cost it also offers additional information about the parametrization, sensitivity and *identifiability* of the approximating model, [26]. Indeed the first and the second derivatives of the rate function $\mathcal{H}(P\,|\,Q^{\theta^n})$ are of the form

$$\nabla_\theta(\mathcal{H}(P\,|\,Q^\theta)) = -\mathbb{E}_\mu\left[\sum_{\sigma'}p(\sigma,\sigma)\nabla_\theta\log q^\theta(\sigma,\sigma')\right]\,, \tag{3.8}$$

and

$$\mathbf{F}_\mathcal{H}\big(Q^\theta\big) = -\mathbb{E}_\mu\left[\sum\sigma' p(\sigma,\sigma)\nabla_\theta^2\log q^\theta(\sigma,\sigma')d\,\sigma'\right]\,. \tag{3.9}$$

The Hessian can be interpreted as a dynamic analogue of the Fisher Information Matrix (FIM) $\mathbf{F}_\mathcal{H}\big(Q^\theta\big)$ on the path space. A similar quantity, in the context of sensitivity

analysis, was recently considered in [26], where the authors also developed efficient statistical estimators for the derivatives of RER $\partial_{\theta_k}\mathcal{H}(P\,|\,Q^\theta)$ and $\partial^2_{\theta_i\theta_j}\mathcal{H}(P\,|\,Q^\theta)$. We discuss related estimators in Section 4.

REMARK 3.1. The proposed approach carries sufficient level of generality in order to be applicable to a wide class of stochastic processes, e.g., Langevin dynamics and KMC, without restriction to the dimension of the system, provided scalable or other efficient simulators are available to simulate the observables, (2.5) and (2.6). The proposed parametrized coarse-graining is applicable to any system for which a *parametrized coarse-grained models* are available, e.g., in coarse-graining of macromolecules and biomembranes, [25, 31, 24]. An obvious obstacle is that the path measure $P_{[0,T]}$ is absolutely continuous with respect to $Q_{[0,T]}$, however, it does not significantly restrict the class of relevant applications as we typically deal with KMC of Markov Chain approximations resulting from a discretization of Molecular Dynamics with noise. In the latter case, Markov chains obtained by numerical approximations of stochastic differential equations (SDEs) allow us to compute RER through (2.5) and can be used for quantification of errors or inverse Monte Carlo fitting for non-equilibrium or irreversible models in Section 3. For example, the diffusion process given by the stochastic differential equations with a d-dimensional Wiener process $W(t)$,

$$dX(t) = b(X(t)) + \sqrt{\epsilon}\, dW(t)\,,$$

can be discretized by the Euler scheme with the time-step h

$$X^{n+1} = X^n + b(X^n)h + \sqrt{\epsilon}\zeta^n\sqrt{h}\,,$$

where $\zeta^n \sim N(0,1)$. In turn, the scheme is equivalent to a Markov chain with the transition kernel

$$p(x,x')dx \sim e^{-\frac{1}{2\epsilon h}|x'-x+hb(x)|^2}\,dx'\,.$$

The parametrization scheme in [10] in the context of Fokker-Planck equations is such an example for SDEs; this scheme is mathematically justified by using the Relative Entropy Rate (2.5) and it is a specific, but reversible (i.e. it has a Gibbs steady state) example of our methodology.

3.2. Path-space likelihood methods and data-based parametrization of coarse-grained dynamics. A different, and asymptotically equivalent perspective on parametrizing coarse-grained dynamics relies on viewing the microscopic simulator as means of producing statistical data in the form of a time-series. Although the proposed method can be applied to systems simulated by Langevin-type dynamics we demonstrate its application in the Kinetic Monte Carlo algorithms in Section 5. The novelty of the presented work lies in deriving the parametrization by optimizing the information content in the path space compared to the available data, taking advantage of computable formulas for relative entropy discussed in Section 2, and by (b) using systematically derived, e.g., via cluster expansions, [17], classes of parametric models giving rise to both statistically identifiable and accurate coarse-grained models. For the latter point we also refer to the discussion of the example in Section 5.

We consider a fine-scale data set of configurations $\mathcal{D} = \{\sigma_1, \sigma_2, ..., \sigma_N\}$ obtained from the KMC algorithm. As is typical in the KMC framework, we assume that the atomistic model, and the corresponding data set, can be described by a spatial, continuous-time Markov jump process, [29]. The path-space measure of this KMC

process, see for the Markov Chain analogue of the path measure (2.3), is parametrized as $P = P^\theta$. In this sense we assume that for the particular data set $\mathcal{D}$ the "true" parameter value is $\theta = \theta^*$. Identifying θ^* amounts, mathematically, to minimizing the pseudo-distance given by the relative entropy, $\min_\theta \mathcal{R}(P^{\theta^*}|Q^\theta)$. Furthermore, following (2.4) it suffices to minimize $\mathcal{H}(P^{\theta^*}\,|\,Q^\theta)$. On the other hand, using the ergodicity of the fine scale process associated with the data set $\mathcal{D} = \{\sigma_1, \sigma_2, ..., \sigma_N\}$, we have the estimators

$$\mathcal{H}(P^{\theta^*}\,|\,Q^\theta) = \lim_{N\to\infty} \hat{\mathcal{H}}_N(P^{\theta^*}\,|\,Q^\theta) \tag{3.10}$$

where we define the unbiased estimator for RER, see Section 4,

$$\hat{\mathcal{H}}_N(P^{\theta^*}\,|\,Q^\theta) := \frac{1}{N}\sum_{i=1}^{N} \log \frac{p^{\theta^*}(\sigma_i, \sigma_{i+1})}{q^\theta(\sigma_i, \sigma_{i+1})}\,, \tag{3.11}$$

and $q^\theta(\sigma, \sigma')$ is defined in (3.3). For simplicity in notation we show the estimator (3.10) for Markov Chain case, where $p^\theta(\sigma, \sigma')$ denotes the transition probability. The continuous-time case, which is relevant to the coarse-grained KMC simulations in Section 5, is similar using (2.6).

Therefore, the minimization of relative entropy rate becomes

$$\min_\theta \hat{\mathcal{H}}_N(P^{\theta^*}\,|\,Q^\theta) = \max_\theta \frac{1}{N}\sum_{i=1}^{N} \log q^\theta(\sigma_i, \sigma_{i+1}) - \frac{1}{N}\sum_{i=1}^{N} \log p^{\theta^*}(\sigma_i, \sigma_{i+1})\,, \tag{3.12}$$

which does not require a priori the knowledge of θ^*. We define the path space Likelihood as

$$L(\theta; \{\sigma_i\}_{i=0}^N) := \frac{1}{N}\sum_{i=1}^{N} \log q^\theta(\sigma_i, \sigma_{i+1})\,. \tag{3.13}$$

Note that if the transition probabilities in (3.13) are replaced with a stationary measure and N corresponding independent samples $\mathcal{D} = \{\sigma_1, \sigma_2, ..., \sigma_N\}$, then (3.13) becomes the classical Maximum Likelihood Principle (MLE). In this sense (3.13) is a Maximum Likelihood for the stationary time series $\mathcal{D} = \{\sigma_1, \sigma_2, ..., \sigma_N\}$ of the microscopic process, and thus include dynamics information. Furthermore, it allows us to obtain the Markovian best-fit from the dynamical simulation and observations on a single, long-time realization of the process.

Fisher Information Matrix and Confidence Intervals. The computable Fisher Information Matrix (FIM) in (3.9) can provide confidence intervals for the corresponding estimator $\hat{\theta}_N \approx \theta^*$, based on the *asymptotic normality* of the MLE estimator $\hat{\theta}_N$. Indeed, under additional mild hypotheses on the samples $\mathcal{D} = \{\sigma_1, \sigma_2, ..., \sigma_N\}$, [8], this general procedure guarantees convergence in the usual central limit sense

$$\hat{\theta}_N \to \theta^* \text{ a.s. and } N^{1/2}(\hat{\theta}_N - \theta^*) \rightharpoonup N(0, \mathbf{F}_{\mathcal{H}}{}^{-1}(Q^{\theta^*})), \tag{3.14}$$

where the variance is determined by the Fisher Information Matrix $\mathbf{F}_{\mathcal{H}}\big(Q^{\theta^*}\big)$, or asymptotically by $\mathbf{F}_{\mathcal{H}}\big(Q^{\hat{\theta}_N}\big)$. Thus estimating the FIM $\mathbf{F}_{\mathcal{H}}\big(Q^{\hat{\theta}_N}\big)$ using (3.9) provides error bars on computed optimal parameter values θ^*.

4. Statistical estimators for RER and FIM. The Relative Entropy Rate (3.5), as well as the Fisher Information Matrix (3.9) are observables of the stochastic process and can be estimated as ergodic averages. Thus, both observables are computationally tractable since they depend only on the local transition quantities. We give explicit formulas for the case of the continuous-time Markov chain and also refer to [26].

The first estimator for RER is given by

$$\widehat{\mathcal{H}}_1^{(n)}(P\,|\,Q^\theta) = \frac{1}{T}\sum_{i=0}^{n-1}\Delta\tau_i\Big[\sum_{\sigma'\in E} c(\sigma_i,\sigma')\times\log\frac{c(\sigma_i,\sigma')}{c^\theta(\sigma_i,\sigma')} - \big(\lambda(\sigma_i)-\lambda^\theta(\sigma_i)\big)\Big]\,, \quad (4.1)$$

where $\Delta\tau_i$ is an exponential random variable with parameter $\lambda(\sigma_i)$ while $T=\sum_i\Delta\tau_i$ is the total simulation time. The sequence $\{\sigma_i\}_{i=0}^n$ is the embedded Markov chain with transition probabilities $p(\sigma_i,\sigma') = \frac{c(\sigma_i,\sigma')}{\lambda(\sigma_i)}$ at the step i and $c^\theta(\sigma_i,\sigma')$ are the rates of the parametrized process, e.g., the coarse-grained rates $c^\theta(\mathbf{T}\sigma_i,\sigma')$. Notice that the weight $\Delta\tau_i$ which is the waiting time at the state σ_i at each step is necessary for the correct estimation of the observable, [11]. Similarly, the estimator for the FIM is

$$\widehat{\mathbf{F}}_1^{(n)} = \frac{1}{T}\sum_{i=0}^{n-1}\Delta\tau_i\sum_{\sigma'\in E} c^\theta(\sigma_i,\sigma')\nabla_\theta\log c^\theta(\sigma_i,\sigma')\nabla_\theta\log c^\theta(\sigma_i,\sigma')^T\,. \quad (4.2)$$

The computation of the local transition rates $c(\sigma_i,\sigma')$ for all $\sigma'\in E$ is needed for the simulation of the jump Markov process when Monte Carlo methods such as stochastic simulation algorithm (SSA), [11] is utilized. Thus, the estimators $\widehat{\mathcal{H}}_1^{(n)}$ and $\widehat{\mathbf{F}}_1^{(n)}$ present only a minor additional computational cost in the simulation.

The second numerical estimator for RER is based on the Girsanov representation of the Radon-Nikodym derivative and it is given by

$$\widehat{\mathcal{H}}_2^{(n)}(P\,|\,Q^\theta) = \frac{1}{n}\sum_{i=0}^{n-1}\log\frac{c(\sigma_i,\sigma_{i+1})}{c^\theta(\sigma_i,\sigma_{i+1})} - \frac{1}{T}\sum_{i=0}^{n-1}\Delta\tau_i\big(\lambda(\sigma_i)-\lambda^\theta(\sigma_i)\big)\,. \quad (4.3)$$

Similarly we can construct an FIM estimator. The term in (4.3) involving logarithms should not be weighted since the counting measure is approximated with this estimator. Unfortunately, the estimator (4.3) has the same computational cost as (4.1) due to the need for the computation of the total rate which is the sum of the local transition rates. Furthermore, in terms of the variance, the latter estimator has worse performance due to the discarded sum over the states σ'. For more details we also refer to [26].

5. Benchmark: coarse-grained driven Arrhenius diffusion of interacting particles.. We demonstrate the proposed methodology on an example of a driven, *non-equilibrium* diffusion process of interacting particles formulated as a lattice gas model with spin variables $\sigma(x)\in\{0,1\}$ at lattice sites $x\in\Lambda_N$.

This is a prototype driven system introduced as a model system for the influence of microscopic dynamics to macroscopic behavior in separations problems in [32]. This model problem is intimately related to works on the structure of non-equilibrium steady states (NESS),[9, 27], as well as to the general formalism of non-equilibrium statistical mechanics, [4, 12]. The evolution of particles is described in the context

of the lattice-gas model as an exchange dynamics with the Arrhenius migration rate from the site $x \in \Lambda_N$ to the nearest-neighbor sites $|y - x| = 1$

$$c(x, y, \sigma) = d\, e^{-\beta(U(x,\sigma))}[\sigma(x)(1 - \sigma(x + 1)) + \sigma(x)(1 - \sigma(x - 1))]\,,$$

which describes the diffusion of a particle at x moving to y and interacting through a two-body potential $J(x - y)$ and with an external field h defining an energy barrier $U(x, \sigma) = \sum_{z \neq x} J(x - z)\sigma(z) - h$. The continuous-time Markov chain is defined by its rates and updates to new configurations $\sigma^{x,y}$ in which the spin variables $\sigma(x)$ and $\sigma(y)$ exchanged its values.

Under the assumption of *a local equilibrium* a straightforward local averaging yields the coarse-grained rates, [19],

$$\bar{c}(k, l, \eta) = \frac{1}{q}\eta(k)(q - \eta(l))d\, e^{-\beta \bar{U}(k,\eta)}\,, \tag{5.1}$$

for the lattice-gas model with local concentrations $\eta(k)$ defined as the number of particles in a coarse cell of the size q; in fact, according to (3.1) we define the coarse graining operator

$$\eta(k) = \mathbf{T}\sigma(k) = \sum_{x \in C_k} \sigma(x)\,, \tag{5.2}$$

where $\eta(k) \in \{0, \ldots, q\}$. Keeping the two-body interactions as a basis for the coarse-grained approximation the effective potentials between block spins k and l are obtained by a straightforward spatial averaging

$$\bar{J}(k, l) = \frac{1}{q^2} \sum_{x \in C_k} \sum_{y \in C_l} J(x - y)\,,$$
$$\bar{J}(k, k) = \frac{1}{q(q - 1)} \sum_{x \in C_k} \sum_{y \in C_k} J(x - y)\,.$$

Assuming that the approximating dynamics is of Arrhenius type we obtain the energy barriers

$$\bar{U}(k, \eta) = \sum_l \bar{J}(k, l)\eta(k) + \bar{J}(0, 0)(\eta(k) - 1) - \bar{h}\,.$$

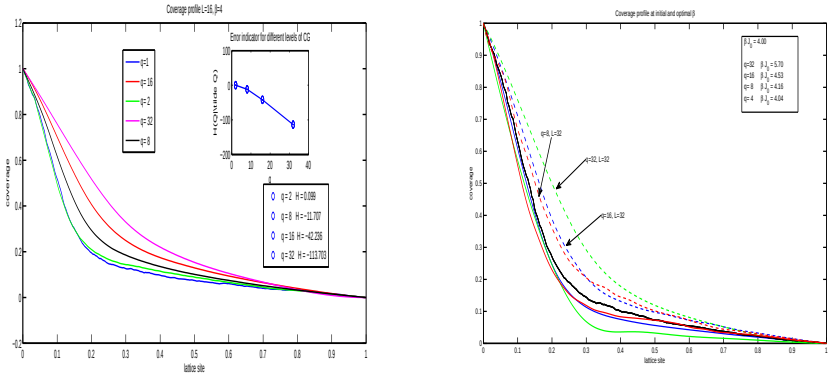
The resulting dynamics is a Markovian approximation of the coarse-grained evolution and it is defined as CTMC with the rates $\bar{c}(k, l, \eta)$.

As a prototype example of the interactions we consider the constant potential $J(x) = J_0$ for $|x| \leq L$ and $J(x) = 0$ otherwise. The effective potential is parametrized by a single parameter $\theta \equiv \bar{J}_0$ corresponding to the strength of the coarse-grained interactions. The system is driven by the concentration gradient given by different concentrations at the boundary sites $x = 0$ and $x = N$. In the long time behavior the distribution converges to a stationary distribution that gives rise to a NESS concentration profile across the computational domain.

The local mean-field approximation which defines the interaction potential $\bar{J}(k-l)$ between two block spins $\eta(k)$ and $\eta(l)$ by averaging contributions from all spin-spin interactions in the cells does not provide a good approximation as demonstrated in Figure 5.1(a), where the inset depicts the error estimated in terms of the entropy rate $\mathcal{H}(P \,|\, \tilde{P})$. However, the mean-field potential $\bar{J}$ is a good initial datum for (3.6).

In this benchmark we stay in the family of two-body potentials and chose to fit only a single parameter that defines the total strength of the interaction. Thus the rates are parametrized by the effective potential $\bar{J}(\cdot;\theta)$ using a single parameter only. The best-fit was obtained by solving the minimization problem (3.4), hence, minimizing the error defined by $\mathcal{H}(P\,|\,\tilde{P}^{\theta})$. Figure 5.1(b) depicts concentration profiles for different sizes q of the coarse cells. The dashed lines represent results from simulations with mean-field interactions between cells only (i.e., the initial guess in the optimization), while the solid lines represent simulations with the parametrized effective interactions.

Comparison with the profile obtained from the microscopic simulation (the solid black line) clearly indicates that when the coarse-graining size q becomes close to the interaction range L of the microscopic potential J the best-fit in a one-parameter family is *not sufficient* for obtaining good approximation and a better candidate class of models, in this case coarse-grained (CG) dynamics $\bar{c}(k,l,\eta)$, needs to be found for improved parametrization. Indeed, in [1, 17] we showed that coarse grained, *multi-body cluster Hamiltonians* provide such a parametrization. More specifically, in [17] we demonstrated, through rigorous cluster expansions that (typical in the state-of-the-art) two-body CG approximations break down in lower temperatures and/or for short range particle-particle interactions, and additional multi-body CG terms need to be included in the models in order the CG model to capture accurately phase transitions and other physical important properties. Hence, no parametrization can consistently address this issue, unless the the proper class of parametric models is identified first.

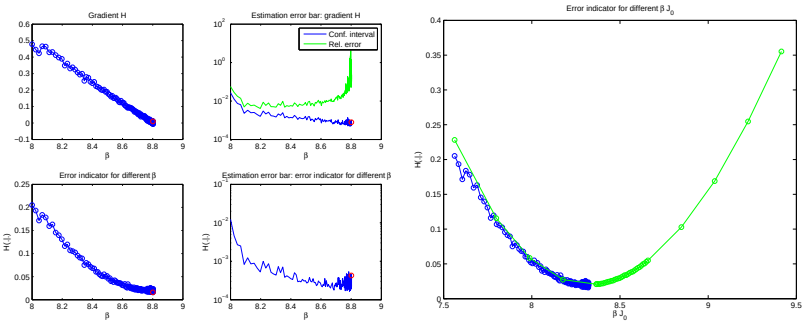


(a) Stationary concentration profiles for different cell sizes q with mean-field interactions. The inset depicts errors at different q estimated by $\mathcal{H}$

(b) Stationary concentration profiles with fitted $\bar{J}(k;\theta^*)$.

FIG. 5.1. *Coarse-grained simulations of driven diffusion of interacting particles without (a) and with (b) fitted effective interactions.*

Acknowledgements. The research of M.A.K. was supported by the National Science Foundation through the CDI -Type II award NSF-CMMI-0835673 and by the European Union (European Social Fund) and Greece (National Strategic Reference Framework), under the THALES Program, grant AMOSICSS. The research of P.P. was partially supported by the National Science Foundation under the CDI -Type



(a) Convergence of estimates $\hat{\theta}_n$ to the optimal value θ^*.

(b) Dependence of $\mathcal{H}$ on the parameter θ.

FIG. 5.2. *Minimization of $\mathcal{H}$. The figure (a) depicts convergence of the estimators for $\hat{\theta}_n$ and the gradient (derivative) $\nabla_\theta \mathcal{H}$ to the optimal value θ^* and the optimality condition $\nabla_\theta \mathcal{H}(\theta^*) = 0$. The right plots depict convergence of confidence intervals for the estimators. The figure (b) demonstrates the convexity of $\mathcal{H}$ with respect to θ which holds due to the particular choice $\theta \equiv \beta \bar{J}_0$ in the benchmark.*

II award NSF-CMMI-0835582. M.A.K. also acknowledges numerous discussions with Yannis Pantazis.

REFERENCES

[1] S. Are, M. A. Katsoulakis, P. Plecháč, and L. Rey-Bellet, *Multibody interactions in coarse-graining schemes for extended systems*, SIAM J. Sci. Comput., 31 (2008), pp. 987–1015.

[2] M. Arnst and R. Ghanem, *Probabilistic equivalence and stochastic model reduction in multiscale analysis*, Comp. methods in applied mech. and eng., 197 (2008), pp. 3584–3592.

[3] I. Bilionis and N. Zabaras, *A stochastic optimization approach to coarse-graining using a relative-entropy framework*, tech. rep., Cornell University, 2012.

[4] F. Bonetto, J. L. Lebowitz, and L. Rey-Bellet, *Fourier's law: a challenge to theorists*, in Mathematical physics 2000, Imp. Coll. Press, London, 2000, pp. 128–150.

[5] S. P. Carmichael and M. S. Shell, *A New Multiscale Algorithm and Its Application to Coarse-Grained Peptide Models for Self-Assembly*, J. Phys. Chem. B, 116 (2012), p. 8383.

[6] A. Chaimovich and M. S. Shell, *Coarse-graining errors and numerical optimization using a relative entropy framework*, 134 (2011), pp. 094112–094112–15.

[7] T. Cover and J. Thomas, *Elements of Information Theory*, John Wiley & Sons, 1991.

[8] M. J. Crowder, *Maximum likelihood estimation for dependent observations*, J. of the Royal Stat. Soc. Ser. B, 38 (1976), pp. 45–53.

[9] J.-P. Eckmann, C.-A. Pillet, and L. Rey-Bellet, *Non-equilibrium statistical mechanics of anharmonic chains coupled to two heat baths at different temperatures*, Comm. Math. Phys., 201 (1999), pp. 657–697.

[10] P. Espanol and I. Zuniga, *Obtaining fully dynamic coarse-grained models from MD*, Phys. Chem. and Chem. Phys., 13 (2011), pp. 10538–10545.

[11] D. T. Gillespie, *A general method for numerically simulating the stochastic time evolution of coupled chemical reactions*, J. of Comp. Phys., 22 (1976), pp. 403–434.

[12] V. Jakšić, C.-A. Pillet, and L. Rey-Bellet, *Entropic fluctuations in statistical mechanics: I. Classical dynamical systems*, Nonlinearity, 24 (2011), pp. 699–763.

[13] E. Kalligiannaki, M. A. Katsoulakis, P. Plechac, and D. G. Vlachos, *Multilevel coarse graining and nano-pattern discovery in many particle stochastic systems*, J. Comp. Phys., 231 (2012), pp. 2599–2620.

[14] M. Katsoulakis, P. Plecháč, L. Rey-Bellet, and D. Tsagkarogiannis, *Coarse-graining schemes and a posteriori error estimates for stochastic lattice systems*, ESAIM: Math Model. Num. Anal., 41 (2007), pp. 627–660.

[15] M. A. Katsoulakis, A. Majda, and D. Vlachos, *Coarse-grained stochastic processes for*

microscopic lattice systems, Proc. Natl. Acad. Sci, 100 (2003), pp. 782–787.

[16] M. A. Katsoulakis, L. Rey-Bellet, P. Plecháč, and D. K.Tsagkarogiannis, *Mathematical strategies in the coarse-graining of extensive systems: error quantification and adaptivity*, J. Non Newt. Fluid Mech., (2008).

[17] M. A. Katsoulakis, L. Rey-Bellet, P. Plecháč, and D. K. Tsagkarogiannis, *Coarse-graining schemes for stochastic lattice systems with short and long range interactions (to appear)*, Math. Comp., (2013).

[18] M. A. Katsoulakis and J. Trashorras, *Information loss in coarse-graining of stochastic particle dynamics*, J. Stat. Phys., 122 (2006), pp. 115–135.

[19] M. A. Katsoulakis and D. G. Vlachos, *Coarse-grained stochastic processes and kinetic Monte Carlo simulators for the diffusion of interacting patricles*, Journal of Chemical Physics, 119 (2003), pp. 9412–9427.

[20] C. Kipnis and C. Landim, *Scaling Limits of Interacting Particle Systems*, Springer-Verlag, 1999.

[21] A. P. Lyubartsev, M. Karttunen, P. Vattulainen, and A. Laaksonen, *On coarse-graining by the inverse Monte Carlo method: Dissipative particle dynamics simulations made to a precise tool in soft matter modeling*, Soft Materials, 1 (2003), pp. 121–137.

[22] A. J. Majda and B. Gershgorin, *Quantifying uncertainty in climate change science through empirical information theory*, Proc. of the National Academy of Sciences, 107 (2010), pp. 14958–14963.

[23] ———, *Improving model fidelity and sensitivity for complex systems through empirical information theory*, Proc. of the National Academy of Sciences, 108 (2011), pp. 10044–10049.

[24] F. Müller-Plathe, *Coarse-graining in polymer simulation: from the atomistic to the mesoscale and back*, Chem. Phys. Chem., 3 (2002), p. 754.

[25] J. T. Padding and W. J. Briels, *Systematic coarse-graining of the dynamics of entangled polymer melts: the road from chemistry to rheology*, Journal of Physics: Condensed Matter, 23 (2011), p. 233101.

[26] Y. Pantazis and M. Katsoulakis, *A relative entropy rate method for path space sensitivity analysis of stationary complex stochastic dynamics*, J. Chem. Phys., 138 (2013), p. 054115.

[27] L. Rey-Bellet and L. E. Thomas, *Exponential convergence to non-equilibrium stationary states in classical statistical mechanics*, Comm. Math. Phys., 225 (2002), pp. 305–329.

[28] B. Smit and T. L. M. Maesen, *Molecular simulations of zeolites: Adsorption, diffusion, and shape selectivity*, Chem. Reviews, 108 (2008), pp. 4125–4184.

[29] M. Stamatakis and D. G. Vlachos, *Unraveling the Complexity of Catalytic Reactions via Kinetic Monte Carlo Simulation: Current Status and Frontiers*, ACS Catalysis, 2 (2012), pp. 2648–2663.

[30] J. Trashorras and D. K. Tsagkarogiannis, *From Mesoscale Back to Microscale: Reconstruction Schemes for Coarse-Grained Stochastic Lattice Systems*, SIAM Journal on Numerical Analysis, 48 (2010), pp. 1647–1677.

[31] W. Tschöp, K. Kremer, O. Hahn, J. Batoulis, and T. Bürger, *Simulation of polymer melts. I. Coarse-graining procedure for polycarbonates*, Acta Polym., 49 (1998), p. 61.

[32] D. G. Vlachos and M. A. Katsoulakis, *Mesoscopic Theories for the Diffusion of Interacting Molecules*, Phys. Rev. Lett., 85 (2000), p. 3898.